

University of Groningen

Responsiveness and Validity of 3 Outcome Measures of Motor Function After Stroke Rehabilitation

Middel, Berrie; van Sonderen, Eric

Published in:
Stroke

DOI:
[10.1161/STROKEAHA.110.580365](https://doi.org/10.1161/STROKEAHA.110.580365)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2010

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Middel, B., & van Sonderen, E. (2010). Responsiveness and Validity of 3 Outcome Measures of Motor Function After Stroke Rehabilitation. *Stroke*, 41(7), E463-E464.
<https://doi.org/10.1161/STROKEAHA.110.580365>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Stroke

American Stroke
AssociationSM

JOURNAL OF THE AMERICAN HEART ASSOCIATION

A Division of American
Heart Association



Responsiveness and Validity of 3 Outcome Measures of Motor Function After Stroke Rehabilitation

Berrie Middel and Eric van Sonderen

Stroke 2010;41:e463-e464; originally published online May 13, 2010;

DOI: 10.1161/STROKEAHA.110.580365

Stroke is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75214
Copyright © 2010 American Heart Association. All rights reserved. Print ISSN: 0039-2499. Online
ISSN: 1524-4628

The online version of this article, along with updated information and services, is
located on the World Wide Web at:

<http://stroke.ahajournals.org/cgi/content/full/41/7/e463>

Subscriptions: Information about subscribing to *Stroke* is online at
<http://stroke.ahajournals.org/subscriptions/>

Permissions: Permissions & Rights Desk, Lippincott Williams & Wilkins, a division of Wolters
Kluwer Health, 351 West Camden Street, Baltimore, MD 21202-2436. Phone: 410-528-4050. Fax:
410-528-8550. E-mail:
journalpermissions@lww.com

Reprints: Information about reprints can be found online at
<http://www.lww.com/reprints>

Letters to the Editor

Stroke welcomes Letters to the Editor and will publish them, if suitable, as space permits. They should not exceed 750 words (including references) and may be subject to editing or abridgment. Please submit letters in duplicate, typed double-spaced. Include a fax number for the corresponding author and a completed copyright transfer agreement form (available online at <http://stroke.ahajournals.org> and <http://submit-stroke.ahajournals.org>).

Responsiveness and Validity of 3 Outcome Measures of Motor Function After Stroke Rehabilitation

To the Editor:

With interest we read the recently published letter of Dr Sivan, "Interpreting Effect Size to Estimate Responsiveness of Outcome Measures,"¹ as a response to a paper by Hsieh et al² in which they provided indices of the magnitude of treatment-related intraindividual change assessed with the Fugl-Meyer Assessment (FMA), Action Research Arm Test (ARAT), and Wolf Motor Function Test performance time (WMFT-TIME) and functional ability scores (WMFT-FAS). As an effect size index, Hsieh et al used the method of the so-called standardized response mean (SRM) by which mean change in scores over time is divided by the SD of these change scores (see Formula A).

$$(A) \quad SRM = \frac{\bar{X}_{\text{change}}}{SD_{(X_{\text{change}})}}$$

As Sivan argued in his letter,¹ the interpretation of the magnitude of intraindividual change estimated with a SRM may lead to overestimation or underestimation of treatment-related effects when the widely used thresholds of Cohen³ are used. These thresholds for classification of the magnitude of mean differences were developed with an effect size index based on standardizing these mean differences using the pooled SD (see Formula B). Dunlap et al convincingly argued that only the pooled SD should be used to compute effect size (ES) for correlated designs and concluded that if the SD_{pooled} is corrected for the amount of correlation between the measures, then the ES estimate will be an overestimate of the actual ES.⁴ It is essential for clinical investigators to understand the differences between the SRM and ES in classifying treatment-related change in terms of Cohen's thresholds (ES <0.20 indicating a "trivial" change, ES between ≥0.20 and <0.50 "small," ES of ≥0.50 to <0.80 a moderate, and ES ≥0.80 a large change).³

$$(B) \quad ES_p = \frac{\bar{X}_{\text{change}}}{SD_{(\text{pooled})}}$$

Hsieh et al refer to our earlier work concerning the risk of misclassification of an SRM when using Cohen's thresholds^{5,6} in their response to Sivan's critic. However, their calculation of adjusted ES estimates⁷ is based on a false assumption. Consequently, adjusting ES for the size of the correlation between baseline and follow-up as computed by Hsieh et al⁷ in the Table leads to an ES estimate more than twice the magnitude of the ES computed using the SD_{pooled} when the correlation between the baseline and follow-up scores is at least 0.8.⁴

Adjustment of a SRM to ES comprises 2 components. First, Cohen introduced a $\sqrt{2}$ correction as necessary for an appropriate use of his tables for sample size calculation. This correction for looking in Cohen's power tables is necessary because these assume 2(N-1) degrees of freedom (2 independent samples), whereas in, for example, pre-/posttest evaluation, only n-1 are actually available³ (pp 46 to 48). Thus, following Cohen's theory, "multiplying SRM by $\sqrt{2}$ (approximately 1.41) compensates for the sample size tables' assumption of double the error variance"³ (p 46).

Second, because the *t* test prescribed in "own control" study designs (baseline to follow-up) is based on correlated means³ (p 48), we also have to compensate for the correlation (*r*) between paired observations. Therefore, according to Cohen, the relative size of the standardizing unit for the SRM to the ES_{pooled} is not

$$(C) \quad d = \frac{d'}{\sqrt{1-r}}$$

but

$$(D) \quad \sqrt{2(1-r)}$$

Thus, the difference between means for paired (dependent) samples needs to be standardized by a value "which is $\sqrt{2(1-r)}$ as large as would be the case were they independent"³ (p 49).

As was shown in an earlier publication, $(d'/\sqrt{2})/\sqrt{(1-r)}$ is equivalent to the SRM and alternatively $SRM * \sqrt{2} * \sqrt{(1-r)}$

Table. Effect Size Estimates SRM, ESP, and ES as Adjusted by Hsieh et al

| Scale | <i>r</i> | <i>r</i> ² | [A] Mean Change | [B] SD_{change} | [C] SD_{pooled} | A/B Effect Size <i>d</i> (SRM) | A/C Effect Size _{pooled} (ESP) | Adjusted Effect Size ⁷ (ESP)/ $\sqrt{(1-r)}$ |
|-----------|----------|-----------------------|-----------------|--------------------------|--------------------------|--------------------------------|-----------------------------------------|---------------------------------------------------------|
| FMA | 0.901 | 0.81 | 5.75 | 4.06 | 8.67 | 1.42 | 0.66 | 2.06 |
| ARAT | 0.915 | 0.84 | 4.68 | 4.95 | 11.24 | 0.95 | 0.41 | 1.40 |
| WMFT-TIME | 0.594 | 0.35 | 2.13 | 5.56 | 5.91 | 0.38 | 0.36 | 0.56 |
| WMFT-FAS | 0.951 | 0.90 | 0.30 | 0.23 | 0.73 | 1.30 | 0.41 | 1.85 |

FMA indicates Fugl-Meyer Assessment; ARAT, Action Research Arm Test; WMFT, Wolf Motor Function Test performance time (WMFT-TIME) and functional ability scores (WMFT-FAS).

(Stroke. 2010;41:e463-e464.)

© 2010 American Heart Association, Inc.

Stroke is available at <http://stroke.ahajournals.org>

DOI: 10.1161/STROKEAHA.110.580365

is equivalent to d' and both indices will vary with the size of r . Hsieh et al have adjusted their ES but may have overlooked the fact that their method of adjustment of SRM to an ES is incomplete. As shown in the Table, correct transformation of their SRM estimates into ES suitable for Cohen's classification resulted in a moderate effect for Fugl-Meyer Assessment and small effects for Action Research Arm Test, Wolf Motor Function Test–TIME, and Wolf Motor Function Test–FAS, respectively, which seems in sharp contrast with the conclusion of Hsieh et al. Extreme high correlations between baseline and follow-up (except Wolf Motor Function Test–TIME) are mainly responsible for overestimation of effect size by the method of Hsieh et al. Minor deviations from published indices are due to calculations without individual data.

Disclosures

None.

Berrie Middel, PhD

University Medical Center Groningen
Department of Health Sciences
Subdivision of Public Health
Department of Oral Sciences and Clinical Epidemiology
University of Groningen
Groningen, The Netherlands

Eric van Sonderen, PhD

University Medical Center Groningen
Department of Health Sciences
Subdivision of Health Psychology
University of Groningen
Groningen, The Netherlands

1. Sivan M. Interpreting effect size to estimate responsiveness of outcome measures. *Stroke*. 2009;40:e709–e711.
2. Hsieh YW, Wu CY, Lin KC, Chang YF, Chen CL, Liu JS. Responsiveness and validity of three outcome measures of motor function after stroke rehabilitation. *Stroke*. 2009;40:1386–1391.
3. Cohen J. *The T Test for Means. Statistical Power Analysis for the Behavioural Sciences*, 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988:19–74.
4. Dunlap WP, Cortina JM, Vaslow JB, Burke MJ. Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*. 1996;1:170–177.
5. Middel B, Van Sonderen FLP. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *International Journal of Integrated Care*. 2002;2:1–21.
6. Middel B, Van Sonderen FLP. Erratum. Statistical significant change versus relevant or important change in (quasi) experimental design: some conceptual and methodological problems in estimating magnitude of intervention-related change in health services research. *International Journal of Integrated Care*. 2008;8:1–2.
7. Hsieh YW, Wu CY, Lin KC. Response to Letter by Sivan. *Stroke*. In press.